

방송 자막을 이용한 인터넷 기반 주요장면 보기 서비스

*김규석 윤정현 김천석 지미경

(주)우리씨에스티

*kangbeko@paran.com

Internet based highlight scene viewing service using DTV closed-caption data

*Kyuseok Kim Jeong Hyun Yoon Cheonseog Kim Mikyong Ji

WooriCST Inc.

요약

VOD(Video On Demand) 서비스는 통·방 융합 환경에서의 킬러서비스로, 지상파 방송사에서는 인터넷을 통해 ‘다시 보기’, ‘TV VOD’ 등의 서비스를 제공하고 있다. 이러한 기존의 방송 VOD 서비스는 콘텐츠를 실제 재생하기 전에는 사용자가 원하는 장면을 찾기가 쉽지 않으며, 또한 콘텐츠를 실제 재생하더라도 대사 내용이 텍스트로 제공되지 않으므로 청각 장애자의 경우 내용을 파악하기가 어렵다.

본 논문에서는 이와 같은 기존의 방송 VOD 서비스의 단점을 보완하기 위해 디지털 방송 프로그램으로부터 자막 데이터를 추출하고 이를 기반으로 방송 콘텐츠의 구간 정보를 생성하기 위한 연구 내용과 이를 적용하여 웹사이트를 통해 제공되고 있는 ‘주요장면 보기’ 서비스의 실 구축 사례를 소개한다.

1. 서론

VOD(Video On Demand) 서비스는 인터넷 등의 통신 회선을 사용하여 원하는 시간에 원하는 영상을 제공/소비할 수 있도록 하는 통·방 융합 환경에서의 킬러서비스로, 국내 지상파 방송사마다 인터넷 웹사이트에서 ‘다시 보기’, ‘TV VOD’ 등의 여러 가지 서비스를 제공하고 있다. 이러한 방송 VOD 서비스는 스트리밍, 다운로드 등과 함께 대본 파일 등을 제공한다.[1] 사용자는 제공되는 요약 정보, 대본 내용 등을 확인하여 시청을 원하는 콘텐츠를 찾을 수는 있으나, 콘텐츠를 실제 재생하여 탐색하기 전에는 시청을 원하는 장면을 찾기가 쉽지 않은 문제가 있으며, 또한 청각 장애인의 경우 콘텐츠를 재생하면서 현재 내용에 대한 대사를 확인할 수 있도록 텍스트가 제공되지 않으므로 내용을 파악하기 어려운 문제가 있다.

방송통신위원회는 2007년 4월 제정된 ‘장애인 차별 금지 및 권리 구제 등에 관한 법률’에 따라 장애인, 노약자, 외국인 등 정보 소외 계층의 방송 접근권 확대를 통한 정보 격차(digital divide) 해소를 목적으로 2008년 4월부터 자막 방송 서비스를 의무화하고 있으며, 2007년 6월 완료된 자막방송 표준[2]에 따라 지상파 방송사에서는 자막 방송 서비스를 제공하고 있다. 현재 국내 지상파 디지털 방송 시스템의 전송 규격인 MPEG-2 TS(Transport Stream)에는 방송 프로그램의 비디오, 오디오 신호 및 다양한 메타데이터가 다중화되어 있으며, 여기에 자막 데이터도 포함되어 있다.

본 논문에서는 MPEG-2 TS로부터 개선된 주요장면 보기 서비스를 위해 필요한 구간 정보를 생성하는 방법을 제안한다.

먼저 디지털 방송 프로그램을 녹화/저장한 MPEG-2 TS 파일을 분석하여 자막 데이터 및 싱크 타이밍을 추출하여 자막 파일을 생성한다.[3] 녹화/저장된 파일에는 일반적으로 프로그램 본 내용 앞 부분에 광고가 포함되므로, 이를 제외하기 위해 시청 등급 안내를 위한 등급

고지 영상, 방송사 표시 로고 화면의 영상을 종합적으로 분석하여 프로그램 본 내용의 실제 시작 지점을 검출[4]하고, 검출된 시간을 기준으로 자막 대사의 발생 시간을 보정한다. 보정된 자막의 대사에 포함된 단어 수 및 시간 정보를 분석하여 여러 개의 구간으로 분할한다.[3] 분할된 구간별로 자막 파일 및 다양한 화면비 및 해상도에 따른 주요 이미지 파일을 생성하고, 이 구간 정보들을 서비스를 위한 콘텐츠 관리 서버에 전송한다.

상기 내용을 본문에서 자막 데이터 추출, 시작점 검출을 위한 영상 분석, 구간 분할 및 구간 정보 생성 항목으로 나누어 상술하고, 마지막으로 서비스 적용 사례를 제시하여 이를 통해 연구 내용의 유용성을 확인한다.

2. 자막 데이터 추출

디지털 방송 서비스가 시작되면서 방송 스트림으로부터 자막 데이터를 추출하기 위한 연구[3]가 진행되었으며 관련 방송 표준[2]도 이미 정의되어있다. 여기서는 입력된 MPEG-2 TS로부터 자막 파일을 생성하기까지의 주요 내용을 소개한다.

입력 MPEG-2 TS에 대해 역다중화를 수행하여 PES(Packetized Elementary Stream)로 구성된 비디오 데이터를 얻어낸다. PES 내에 사용자 데이터(Picture user data)가 포함되어 있으며, 이 안에 자막 데이터를 넣도록 규정된 cc_data와 cc_count 필드로부터 자막 데이터를 구성한다.

추출된 자막 데이터를 기반으로 SAMI(Synchronized Accessible Media Interchange) 규격을 준수하는 자막 파일을 생성하기 위해서는 자막 데이터가 발생하는 시간, 즉 동기화 시간 정보가 필요하므로 PES의 헤더에 포함되어 있는 33bit의 PTS(Presentation Time Stamp) 값을 변환하여 동기화 시간을 구한다.

이와 같이 구한 자막 데이터와 동기화 시간을 HTML 형식의 텍스트로 기록하여 SAMI 자막 파일을 생성한다.

3. 시작점 검출을 위한 영상 분석

방송사에서 송출된 디지털 방송 프로그램을 EPG 정보를 이용하여 예약 녹화를 하게 되면 저장된 파일에는 일반적으로 프로그램 본 내용과 함께 그 앞 부분에 광고가 포함된다. 2절에서 생성된 자막 파일의 동기화 시간은 저장된 파일의 시작점을 기준으로 정해지는데, 실제 서비스를 위한 콘텐츠의 시작점은 프로그램 본 내용을 기준으로 정해지므로 광고 구간 길이만큼 보정해야 한다.

이와 같은 처리를 위해 여기서는 프로그램 시작 부분의 특징을 기술하고, 이 특징에 따른 프로그램 본 내용의 시작점 검출을 통해 보정하는 방법을 기술한다.

가. 프로그램 시작 부분의 특징

‘방송프로그램의 등급분류 및 표시 등에 관한 규칙’에 따라 방송 프로그램 본 내용의 시작점 등급기호와 함께 부연설명을 5초 이상 자막고지하도록 규정되어 있다. 방송사에서는 일반적으로 별도의 고지영상을 통해 이 규정을 준수하고 있다. 또한 각 방송사마다 방송 프로그램 본 내용이 시작되면 화면의 일정 위치에 방송사 로고를 표시한다. 그림 1은 등급고지영상의 예, 그림 2는 방송사 로고 표시의 예를 나타낸다.

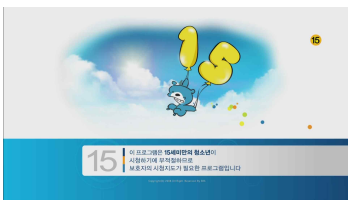


그림 1. 등급고지영상



그림 2. 방송사 로고 표시

나. 등급고지영상 기반 시작점 검출

일반적인 방송 프로그램 영상은 다양한 객체가 포함되며 그 구성 또한 시간이 경과함에 따라 불규칙적으로 바뀐다. 반면 등급고지영상은 약 5초 길이 분량으로 영상의 내용이 고정되어 있다. 이와 같이 일정한 길이를 가지며 고정된 내용을 포함하는 영상을 식별하기 위해서는 영상의 여러 가지 특징 중 컬러와 텍스트 특징을 복합적으로 적용하는 것이 효과적이다.

등급고지영상의 판별을 위해서는 먼저 기준이 되는 영상으로부터 주요 I 프레임 이미지를 추출하고, 이들로부터 컬러와 텍스트의 특징벡터를 추출하여 특징벡터 DB를 구성한다.(그림 3) 광고가 포함된 실제 입력 영상에서 일정 시간마다 주기적으로 이미지를 추출하고, 이로부터 특징벡터를 추출하여 특징벡터 DB와 유사도를 비교한다.(그림 4) 유사도 측정값들 중 최소값이 임계점 이하이면 등급고지영상으로 판별한다.

본 논문의 실험에서는 컬러 특징은 MPEG-7 Color Histogram Descriptor[5]를 응용한 다중 양자화 알고리즘을 적용하였으며, 텍스트 특징은 MPEG-7 Edge Histogram Descriptor[5]를 적용하여 이들을

복합한 특징벡터를 사용하였다.

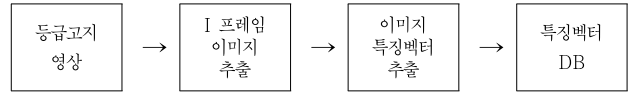


그림 3. 특징벡터 DB 구성

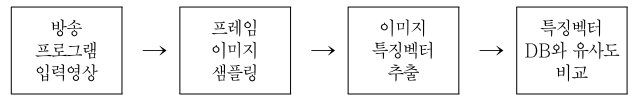


그림 4. 등급고지영상 판별 과정

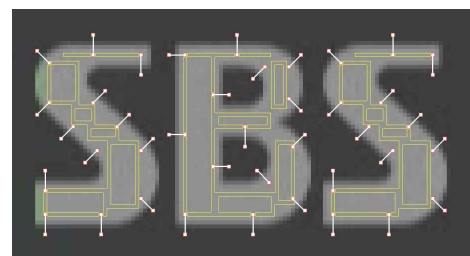
다. 방송사 로고 기반 시작점 검출

방송 프로그램의 장르에 따라 등급고지영상이 생략되는 경우가 있다. 뉴스, 스포츠 중계, 시사 토론 등의 프로그램이 이에 해당하며 이 경우는 3절 나에서 기술한 등급고지영상 판별 알고리즘을 적용할 수 없으므로 프로그램 본 내용에서 나타나는 방송사 로고 여부를 판별한다.

방송사 로고의 형태는 일반적으로 백색 반투명 그래픽으로 처리되어있으며 화면의 좌상단 혹은 우상단에 항상 고정된 위치에서 나타난다. 백색 반투명 그래픽 처리에 의해 로고 영역은 최소 밝기값을 가지므로, 먼저 로고 영역 내에 블록을 설정하고 블록 내 모든 픽셀의 최소 밝기값을 확인한다. 여기서 로고 영역의 가장자리는 그라데이션 처리로 인해 밝기값이 점진적으로 변하기 때문에 그림 5(a)와 같이 이러한 영향을 받지 않는 영역으로 블록을 설정한다. 이러한 최소 밝기값 조건은 로고 유무와 상관없이 해당 영역이 전체적으로 밝은 경우에도 충족되므로, 두 번째 단계로 로고 영역 내부와 외부의 밝기값의 차이를 비교한다. 이를 위해 로고 영역 내부의 픽셀과 인접한 외부 영역의 픽셀을 쌍으로 샘플링하여 밝기값의 차이를 구하고 이를 임계값과 비교한다. 그림 5(b)는 픽셀쌍을 설정한 예시이다. 로고 영역 내부 및 주변의 밝기값이 커질수록 픽셀쌍의 밝기값 차이는 작아지므로 밝기값이 커짐에 따라 판별을 위한 임계값을 작게 적용한다.



(a) 블록 설정



(b) 픽셀쌍 설정

그림 5. SBS 방송사 로고 분석

4. 구간 분할 및 구간 정보 생성

영상의 구간 분할을 위한 다양한 접근 방법을 통한 연구가 진행되어 왔다. 비디오 프레임 분석에 의한 장면 분할 방법에 많은 연구 사례가 있으며, 또한 자막 데이터를 기반으로 하여 시간적으로 분할하는 방법[3]도 연구되었는데, 여기서는 처리 속도에서 매우 우위에 있는 자막 내용 분석을 이용한 구간 분할 방법을 적용하였다.

자막은 프로그램 장르에 따라 구별되는 특징이 있는데, 특히 뉴스의 경우 ‘앵커’, ‘기자’, ‘인터뷰’ 등과 같은 일종의 태그가 대사 앞에 붙어 화자를 구별할 수 있다. ‘기자’ 대사 다음에 ‘앵커’ 대사 나오는 경우 뉴스가 바뀌는 경우가 많으므로 이를 기준으로 구간을 분할할 수 있다.

```
<SVNC Start=79656><P Class=KRCC>-<앵커> 우리 대표팀의 첫 상대인 대만의 요주 인물이
<SVNC Start=82325><P Class=KRCC>있습니다.
<SVNC Start=82492><P Class=KRCC> 두 차례 평가전에서 맹타를 휘두른 3번 타자 권이휘인
<SVNC Start=85992><P Class=KRCC>&nbsp;&nbsp;&nbsp;
<SVNC Start=87831><P Class=KRCC>선수입니다.
<SVNC Start=89499><P Class=KRCC>도쿄에서 임창정 기자가 취재했습니다.
<SVNC Start=91668><P Class=KRCC>
<SVNC Start=94084><P Class=KRCC>-<기자> 한국 대표팀의 권이휘엔 경계령이 내려졌습니다.
<SVNC Start=96172><P Class=KRCC>권이휘엔은 일본 대표팀과 가진 2차례 평가전에서 총런
<SVNC Start=99342><P Class=KRCC>1개를 포함해 10타수 6안타를 득표시켰습니다.
<SVNC Start=102842><P Class=KRCC>&nbsp;&nbsp;&nbsp;
<SVNC Start=103680><P Class=KRCC>장타력뿐만 아니라 정교함과 빠른 발까지 갖춰 우리 투수진을
<SVNC Start=107180><P Class=KRCC>&nbsp;&nbsp;&nbsp;
<SVNC Start=107680><P Class=KRCC> 괴롭힐 전망입니다.
<SVNC Start=111184><P Class=KRCC>&nbsp;&nbsp;&nbsp;
<SVNC Start=117027><P Class=KRCC>
<SVNC Start=117360><P Class=KRCC>-<기자> 오늘 세이브전에서 두 점 출런을 친 2루수 지양즈시엔
<SVNC Start=120860><P Class=KRCC>&nbsp;&nbsp;&nbsp;
<SVNC Start=122699><P Class=KRCC>도 주의해야 할 선수입니다.
<SVNC Start=123867><P Class=KRCC>마운드는 다소 불안합니다.
```

그림 6. 뉴스 프로그램의 자막 예

드라마를 포함한 대부분의 프로그램 장르에서는 청각 장애인 등을 위해 화자 전환 시 자막 데이터의 대사 앞에 하이픈('-') 기호가 삽입되어 있는 특징을 이용할 수 있다. 화자가 전환되어 하이픈('-') 기호가 나타날 경우, 앞 대사와 현재 대사가 각각 발생한 시각을 비교하여 이 시각이 임계값 이상이면 구간을 분할한다. 특히 드라마는 내용 중에 대사가 나오지 않는 샷이 나타나므로 이 부분에서 구간을 분할할 수 있는데, 이러한 샷이 빈번하게 나타나면 길이가 매우 짧은 구간들이 발생하므로 이러한 구간은 이전 또는 다음 구간에 병합한다.

이와 같이 구간을 분할하여 구한 각 구간의 시작 시간과 끝 시간에 따라 구간별로 SAMI 규격의 자막 파일을 생성한다. 또한 각 구간의 내용을 이미지를 통해 미리 확인할 수 있도록 주요 장면 이미지를 생성한다. 방송 프로그램의 MPEG-2 TS 파일로부터 프레임 이미지를 추출하면 원본의 해상도는 1920x1080 크기를 가지는데, 이를 디스플레이선(decimation) 처리하여 인터넷 상에서 서비스할 수 있도록 작은 크기의 썸네일 이미지를 생성한다. 현재 일부 방송 프로그램의 경우 4:3 화면비로 제작되어 영상의 좌우에 레터박스 처리되므로 이미지 생성의 전처리 과정에서 이 부분을 제거한다.

그림 7은 방송 프로그램 입력 파일 MPEG-2 TS로부터 구간 정보를 생성하기까지의 전체 처리 과정을 블록 다이어그램으로 나타낸 것이다.



그림 7. 구간 정보 생성을 위한 전체 처리 과정

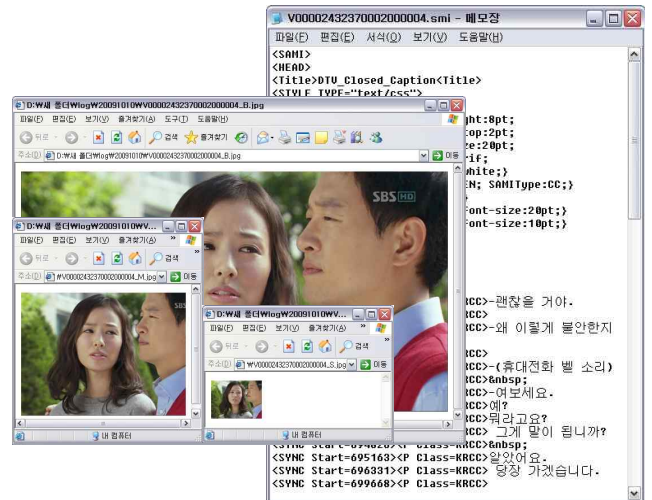
5. 실험 및 적용 사례

가. 실험

그림 8은 그림 7에서 도시한 처리 과정을 구현한 자막 처리 프로그램의 실험 화면이다. 본 실험에서 사용한 영상은 PC의 TV 수신카드를 이용하여 녹화/저장한 MPEG-2 TS 파일이며, 본 실험에서는 이를 자막 처리 프로그램의 입력으로 사용하였다. 그림에서 보이는 입력 영상은 드라마 장르의 프로그램으로 본 내용의 앞 부분에 등급금지영상이 삽입되어 있다.



(a) 구간 분할 처리 화면



(b) 주요 장면 이미지

(c) 구간 자막 파일

그림 8. 자막 처리 프로그램

먼저 MPEG-2 TS 파일을 분석하여 자막 데이터를 구성하고, 각 데이터의 동기화 시간을 추출하여 전체 내용에 대한 자막 파일을 생성한다. 이와 동시에 파일에서 프로그램 본 내용의 시작점을 검출하기 위해 다음의 과정을 반복하여 수행한다. 영상의 처음을 기준으로 프레임 이미지를 주기적으로 추출하고 각 이미지에 대한 특징벡터를 추출한다. 추출된 특징벡터를 등급금지영상의 특징벡터 DB와 유사도를 비교하여 등급금지영상 여부를 판별한다.

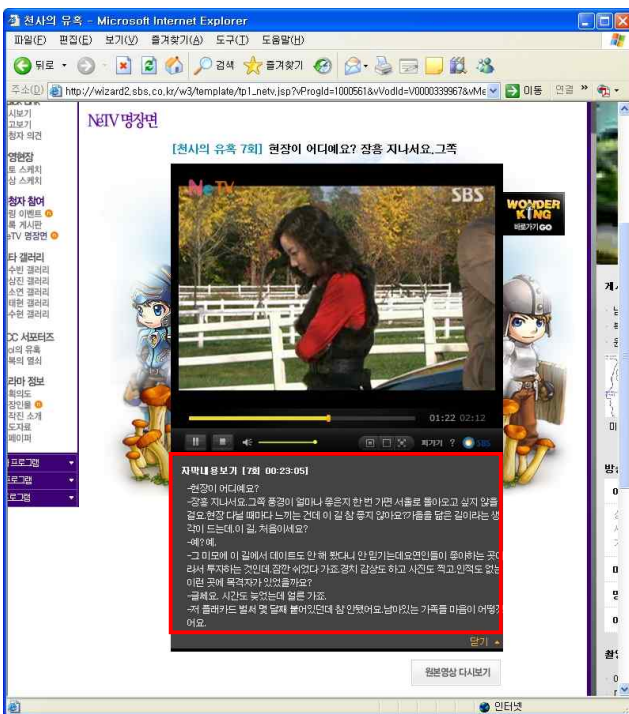
이 과정을 거쳐 등급고지영상으로 최초로 판별된 프레임 이미지의 추출 시간값을 적용하여 전체 자막 파일의 모든 대사의 동기화 시간을 보정함으로써, 광고 부분을 제외한 방송 프로그램 내용에 대한 자막 파일(SAMI)을 생성한다.

다음 단계로 이와 같이 생성된 자막 내용을 분석하여 구간을 분할하고 짧은 구간은 병합하여 구간 분할이 완료되면, 각 구간의 시간 정보에 따라 전체 자막의 내용을 분할하여 구간별로 자막 파일을 저장하고, 각 구간의 첫 장면에서 입력 영상의 원본 프레임 이미지 및 썸네일 이미지를 생성한다.

그림 8(a)는 자막 처리 프로그램을 실행하여 구간 분할이 완료된 화면을 나타낸다. 오른쪽의 목록에서 분할된 구간의 목록을 나타내며, 붉은 사각형으로 표시한 구간의 주요 장면 이미지와 구간 자막 파일을 (b), (c)에서 나타낸다.



(a) 자막영상 메뉴 화면



(b) 구간 재생 화면

그림 9. SBS 웹사이트의 자막영상 서비스

나. 적용 사례

본 논문의 연구를 통해 구현된 자막 처리 기술은 방송사 SBS 웹사이트에서 제공되는 '자막영상' 서비스에 활용되고 있다. 자막 처리 프로그램을 통해 생성된 구간별 자막 파일과 주요 장면 이미지 등의 구간 정보를 SBSi 콘텐츠 관리 서버로 전송함으로써, 이를 이용한 새로운 VOD 서비스가 일반 사용자를 대상으로 제공된다.

그림 9는 현재 SBS 웹사이트에서 '자막영상' 메뉴로 제공 중인 주요장면 보기 서비스로서, (a)는 구간 자막 내용, 주요 장면 이미지 등의 구간 정보를 제공하는 화면이며, (b)는 선택된 구간에 대해 자막과 함께 스트리밍 서비스를 제공하는 화면을 나타낸다.

구간 정보를 이용해 콘텐츠를 실제 재생하여 탐색하기 전에 원하는 장면을 찾을 수 있으며, 또한 청각 장애인의 경우에도 콘텐츠를 재생하면서 구간 자막을 통해 내용을 확인할 수 있음을 알 수 있다.

6. 결론

본 논문에서는 개선된 주요장면 보기 서비스를 위해 필요한 구간 정보를 생성하는 방법을 제안했다.

디지털 방송 프로그램용 포맷인 MPEG-2 TS 파일에서 자막 데이터 및 싱크 타이밍을 추출하여 자막 파일을 생성하고, 영상 분석을 통해 자막의 시작점을 보정된 자막 파일을 생성한다. 자막 데이터를 기반으로 방송 프로그램의 구간을 분할하고 구간 정보를 생성한다.

이 구간 정보를 이용한 방송사의 개선된 주요장면 보기 서비스를 통해, 제안된 자막 처리 기술의 유용함을 확인할 수 있다.

참고문헌

- [1] <http://www.sbs.co.kr>, <http://www.imbc.com> 등
- [2] "지상과 디지털 TV 자막방송", TTA, TTAS. KO-07.0050, Jun. 2007
- [3] "디지털방송 자막데이터를 이용한 멀티미디어 응용 서비스 연구", 김정연 외, 한국방송공학회논문지 제14권 4호, Jul. 2009
- [4] "방송 프로그램의 주요 특징을 이용한 시작점 검출 방법", 윤정현 외, 신호처리합동학술대회 Vol.22, No.1, 2009
- [5] S. Manjunath(Editor), Philippe Salembier(Editor) and Thomas Sikora(Editor): Introduction to MPEG-7: Multimedia Content Description Interface. Wiley&Sons, April 2002